

# ECON832 Final: Mini Paper

Daniel Sánchez Pazmiño

April 2024

## 1 Overview

The data comes from the CPC18 (Plonsky, Erev, and Ert 2017), which consists of experiments involving decision makers choosing between two lotteries. I focus on data exclusive to CPC18 (Experiment 1), which added attention features to the CPC15 data (Erev, Ert, and Plonsky 2017). This experiment involved sixty choice problems for 240 participants. I use the calibration data set to train my deep feedforward neural network and the **individual track data** (Track II) to test the model. All replicating code and files are available on the project's GitHub repository.

## 2 Methodology

### 2.1 Data preparation

The calibration dataset was filtered to only include Experiment 1 data (sets 5 and 6). The outcome variable was the binary indicator of the choice of B over A. Since the variable was already dichotomous, no transformation was needed. I include demographic variables in the baseline models: location, gender and age. I dichotomized categorical variables with Technion and male as the reference levels.

Risk features involve the number of lottery outcomes, the expected values, the low payoffs, and probabilities to draw from the lotteries and lot distribution shapes. The latter defines the distribution of the lotteries, which can be binomial around the expected value, right skewed or left skewed. They were dichotomized in the same way as the location and gender variables.

Other relevant variables involve whether there is ambiguity in the probabilities or if there is a correlation between the payoffs. Further, it is important to include trial numbers, potential and obtained payoffs obtained, and forgone payoffs. Some of these variables are categorical, but they are already dichotomized, so I didn't transform them.

Regarding attention features, I include the number of time block, reaction time, feedback binary outcome, the onscreen button and the trial number within a game.

The processing of the calibration dataset was done using `TidierData.jl` and `DataFrames.jl`. I performed a test-train split with 80% of the data used for training and 20% for testing. I further tested with the individual track data Plonsky, Erev, and Ert (n.d.).

## 2.2 Feature engineering

As mentioned, all categorical variables for both the risk and attention features were dichotomized. This is equivalent to **one-hot encoding**. The continuous variables were **standardized** to have  $\mu = 0$  and  $\sigma = 1$ . The data was processed from its rectangular form (every row per decision) to a format where each row represented an attribute, and every column represented a decision.

## 2.3 Model configuration

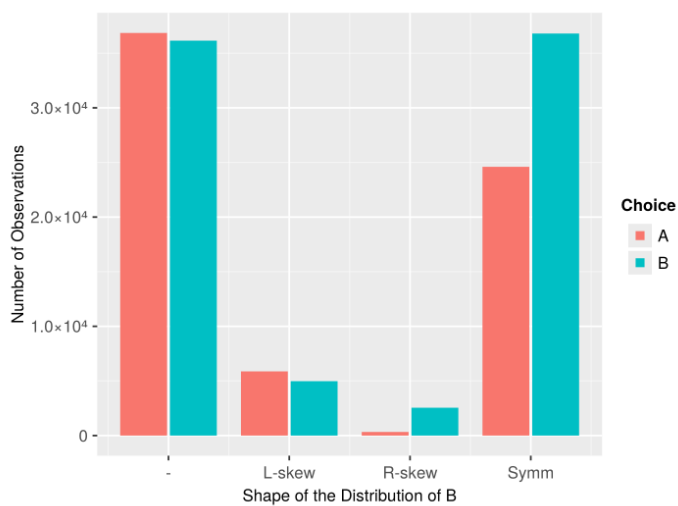
The model is a deep feedforward network neural network (DFNN) with `Flux.jl`. I use a ReLU activation function for the input and hidden layers and a sigmoid activation function for the output layer. I use MSE as the loss function and the gradient descent optimizer to minimize the loss function. The learning rate was set to 0.7, and I train the model for 100 epochs using a `for` loop. I use a confusion matrix and an accuracy rate (percent correctly predicted) to evaluate the model's fit, using a 0.5 probability threshold to determine predicted classes.

# 3 Feature analysis

## 3.1 Risk features

The correlation between the number of outcomes and the choice for the B lottery is 0.09, which is statistically significant at the 1% level. Further, the correlation between the low payoff of the lotteries and the choice for the B lottery is 0.11, which is statistically significant at the 1% level. This might be a result of the risk aversion of the participants, which is a common feature in decision-making and point towards the importance of considering risk-aversion in the model rather than risk neutrality. The correlation between lotteries shows a relationship with the choice of B, with negative correlation of -0.14, which is statistically significant at the 1% level.

Figure 1: Choice of lottery by shape of distribution



In terms of the shape of the distribution, I identify that a symmetric distribution greatly favours the choice of B, as seen in Figure 1. Further, I also identify that the payoff that the participant received shows a statistically significant correlation with the choice of B. I choose to include all 12 variables which jointly define the shape of the distribution in the model, as well as the payoff that the participant received plus the potential payoffs of A and B.

### 3.2 Attention features

## References

- Erev, Ido, Eyal Ert, and Ori Plonsky. 2017. “Raw Data for CPC2015: A Choice Prediction Competition for Decisions Under Risk, Under Ambiguity, and from Experience.” Data set with codebook. Zenodo. <https://doi.org/10.5281/zenodo.321652>.
- Plonsky, Ori, Ido Erev, and Eyal Ert. 2017. “Calibration Data for Choice Prediction Competition 2018 (CPC18).” Data set with codebook. Zenodo. <https://doi.org/10.5281/zenodo.845873>.
- . n.d. “Competition Set Data Track II (Individual Behavior): Raw Data.” Data set. Zenodo. Accessed April 20, 2024. [CPC18: Choice Prediction Competition 2018](#).